DOCUMENT RESUME

ED 178 564 TM 009 697

AUTHOR	Roid, Gale; And Others
TITLE	A Comparison of Methods for Transforming Sentences into Test Questions for Instructional Materials. Technical Report #1.
INSTITUTION	Advanced Research Projects Agency (DOD), Washington, D.C.
SPONS AGENCY	Navy Personnel Research and Development Center, San Diego, Calif.
PUB DATE	28 Feb 78
CONTRACT	MDA-903-77-C-0189
NOTE	50p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Algorithms; *Criterion Referenced Tests; Higher Education; Item Analysis; *Multiple Choice Tests; Prose; *Questioning Techniques; Science Tests; Senior High Schools; Structural Analysis; *Test Construction; *Test Items
IDENTIFIERS	Test Format

ABSTRACT

Several measurement theorists have convincingly argued that methods of writing test questions, particularly for criterion-referenced tests, should be based on operationally defined rules. This study was designed to examine and further refine a method for objectively generating multiple-choice questions for prose instructional materials. Important sentences were selected from a prose passage in a science text and these sentences were transformed into questions. Several variations of sentence transformation rules were used to create tests given to 273 college and high school students before and after they read the passage. Item difficulties (percent correct) for each type of item formed the basic data of the study. The study concluded that the method of selecting the "question word" (a noun or adjective) in the sentence has a crucial role in determining the pattern of pretest and posttest item difficulties of the resulting question. Also, the methods of item writing used in the study were found to be feasible and to be free from the item-writer differences that typically are found in traditional item-writing methods. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED178564

A COMPARISON OF METHODS FOR TRANSFORMING SENTENCES INTO TEST QUESTIONS FOR INSTRUCTIONAL MATERIALS

Gale Roid and Tom Haladyna
Teaching Research Division
Oregon State System of Higher Education

Patrick Finn
State University of New York at Buffalo

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Gale Roid
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Technical Report #1

for

Contract Number MDA 903-77-C-0189
Defense Advanced Research Projects Agency

Submitted to:

Navy Personnel Research and Development Center
San Diego, California 92152

February 28, 1978

TM009 697

FOREWORD

The research and development described in this report is related to ZF55.522.012.03.01, Criterion-referenced Testing/Training, and was conducted under contract number MDA 903-77-C-0189 from the Defense Advanced Research Projects Agency. Technical monitor for the contract has been Dr. Pat-Anthony Federico, Navy Personnel Research and Development Center, San Diego.

Appreciation is expressed to Dr. John R. Bormuth of the University of Chicago and Dr. Jason Millman of Cornell University who have been consultants for the project.

This technical report describes the first phase of research aimed at examining the qualities of test questions written from a variety of methods, and is based on the first in a series of related experiments.

SUMMARY

Problem

Several measurement theorists have argued convincingly that methods of writing test questions, particularly for criterion-referenced testing, should be based on objective, operationally-defined rules. Subjective methods of writing questions ("items"), such as those used to write questions from learning objectives, have the disadvantages that they are (1) not open to public inspection, (2) not capable of being replicated by item writers other than the original test developer, (3) not assured of being precisely related to the instruction, and (4) subject to item writer bias. For these reasons, a technology of item writing is needed. There is a particular need for methods of writing items to test learning from prose passages in instructional materials.

Objective

To examine and further refine a method of subjectively generating multiple choice test questions by transforming sentences from prose instructional materials.

Approach

An algorithm developed by Finn (1975) was used to derive test items from a prose passage on "insect development." There were 273 volunteer college and high school students who served as subjects. The algorithm involves a computer analysis of the prose passage and identifies two kinds of words: (1) "high information" words--nouns or adjectives that are relatively rare in American textbooks and only appear a single time in the passage, and (2) "keywords"--nouns or adjectives that are rare in American textbooks yet occur several times in a given passage. A sample is selected of high information words that are either adjectives or nouns; sentences in which these words appear are transformed into questions by placing a wh-word, such as "who" or "what," in the front of the sentence to form a question. A random sample of words from the passage that are in the same semantic classification as the question word are used for foils in multiple-choice versions of the questions. Items were written by Finn and three graduate students, using this algorithm and several variations including one in which foils were simply selected by each writer. Tests consisting of a mixture of item types were given to subjects before and after reading the insect development passage.

Items, rather than subjects, were the units of analysis for this study. The difficulties (percent correct answers) for each of the 160 items on pretest and posttest were computed and formed the basic data for the study.

Findings

1. Keyword nouns that are relatively rare in American textbooks, but occur frequently in the particular passage were found to produce low quality items when the first occurrence of the noun in the passage was used to form

the question. Pretest difficulty was a high 61.3%, indicating that subjects could easily guess, before instruction, the correct answer to questions in which these nouns occurred.

2. "High information" nouns and adjectives and "keyword" adjectives were found to show a significant change in item difficulty from pretest to posttest, indicating a good quality of item for the type of prose instruction used in this experiment.

3. No significant differences between item writers was found in the study of item difficulties, indicating that the sentence transformation methods employed in this study apparently neutralized the effects of item writer bias that has been found in other studies of item writing.

4. In a study of the variability of item difficulties across item writers, no significant differences in variance were found between "writer's choice" foils and algorithmically generated foils.

Conclusions

The methods of item writing implemented in this study are apparently effective in reducing the item writer bias identified in previous studies by the authors of this report (Roid and Haladyna, 1978).

High information words (rare singletons) appear to be promising candidates to include in questions that test learning from prose. The first occurrence of nouns that are relatively rare in American textbooks but highly frequent in a passage are apparently not good candidates for questions. Algorithmic methods of generating foils for multiple-choice questions that are based on selecting adjectives or nouns from a prose passage are feasible and produce items of reasonably good quality. Quality was assessed by the pattern of item difficulties on pretest and posttest, and the variability of item difficulties across item writers.

Recommendations

1. Algorithmic generation of multiple-choice test questions for assessing rote learning of prose should be considered feasible.

2. "High information" words, as defined in this study, should be considered a useful basis for selecting important sentences from prose passages for transformation into questions.

3. The first occurrence in a passage of "keyword" nouns, as defined in this study, should not be used in generating items from a prose passage.

CONTENTS

	page
INTRODUCTION	1
Problem	1
Background	3
Bormuth's Theory of Achievement Test Items	3
Finn's Extension of the Bormuth Methodology	4
Summary	8
Objectives	9
APPROACH	11
Materials Used in the Study	11
Selection of Question Words	11
Item Writers	12
Development of Foils	12
Items	15
Research Hypothesis	15
Subjects	16
Data Analysis	16
RESULTS AND DISCUSSION	18
Variance Between Writers	24
CONCLUSIONS	27
RECOMMENDATIONS	29
REFERENCES	30
REFERENCE NOTES	32
APPENDIX A - "INSECT DEVELOPMENT"	A-0
APPENDIX B - EXAMPLES OF ITEMS PRODUCED FROM THE PASSAGE "INSECT DEVELOPMENT"	B-0
APPENDIX C - STEPS IN TRANSFORMING SENTENCES INTO TEST QUESTIONS	C-0

TABLES

1. Repeated Measures Analysis of Variance on Item Difficulties of Items of Each Type	19
2. Means and Standard Deviations of Item Difficulties	20
3. Means and Standard Deviations of Item Difficulties for the Noun and Adjective Items on Pretest and Posttest (P and RP Effects)	21

TABLES, Cont'd.

Page

4. Means and Standard Deviations of Item Difficulties for the PS and RPS Interaction Effects 21

5. Keywords in the Insect Development Passage and Their Text Frequency 23

6. Means and Standard Deviations for the RFS Interaction Effect 24

7. Variabilities and Standard Deviations of Item Difficulties 26

INTRODUCTION

Problem

Measurement theorists have convincingly argued in recent years that there is a crisis in education that stems from the lack of a scientific basis for writing achievement test questions ("items"). This crisis has been prompted by the increased public demand for accountability in education and interest in the use of tests for selection, placement, advancement, certification, and other important decisions which affect people's lives. Currently, the test specialist must provide tests that supply information for making important decisions, without a systematic technology of test item writing. A highly developed technology of item writing is not present even in the prominent method of criterion-referenced testing where an individual's performance is compared to a standard rather than to the performance of other individuals.

The most widely used methods of writing test questions, for both criterion-referenced and traditional norm-referenced tests, rely on the intuitive skills of the item writer or panels of experts who judge the merits of questions. Even when item writers are given learning objectives that describe what is to be learned in terms of expected student performance under specified conditions and standards, it has been shown in research, by the author that two writers will not generate the same items or items of similar quality. The current military guidelines for designing criterion-referenced tests for use in instructional systems (TRADOC Pamphlet 350-30; Swezey and Pearlstein, 1974) refer to the "writing of test items for each learning objective," but do not provide detailed suggestions for an item writing technology. A technology of item writing must be developed to eliminate these deficiencies in conventional methods, by providing methods that are:

1. Based on a logical and precisely defined relationship between the instruction given and the test items written.

2. Defined by a set of operations open to public inspection rather than defined by the private, intuitive standards of the item writer.

3. Specific and public so that they can be easily replicated by many test developers, and not just the original item writer.

Item writing methods with these attributes will allow tests to become more scientific instruments, and contribute to the advancement of instructional research, educational evaluation, and the use of test data in the formation of public policy.

Although theories and suggestions have been published concerning new item writing methods, little specific research has been conducted to determine the technical quality of items written by proposed new methods nor the feasibility of their widespread use in education and training. There have been only a handful of civilian research studies, most of which are currently unpublished, that have examined the technical and measurement qualities of the new item writing methods such as those capable of automation. Clearly, these methods should have a strong research base if they are to be used in military training and to reshape the everyday practices of educational testing in the United States.

There is also a practical reason for interest in algorithmic methods of writing test questions. A large number of items are needed, particularly for instructional systems involving the mastery learning model (Bloom, 1968), in order to provide multiple test forms for use by students who may be retested several times. Each of these test forms must be equivalent in terms of content coverage and difficulty. Although test forms could be field tested and their difficulties assessed and revised, a great savings in time and energy could be achieved if test forms of near equivalency could be produced algorithmically prior to use. The challenge of producing many items of good

quality has prompted considerable interest in a technology of test item writing.

Background

Roid and Haladyna (1978) have discovered that Millman (1974) and Bormuth (1970) and others are on firm empirical grounds when they conclude that "two writers using learning objectives will not necessarily write comparable items." The Roid and Haladyna study showed that one of two item writers produced consistently more difficult test items from the same learning objectives. This bias results in tests of different difficulty depending on the writer, which would have serious implication for the criterion-referenced uses of such tests, e.g., affecting decisions about who passes or fails such tests.

Bormuth's Theory of Achievement Test Items

Bormuth (1970) proposed operationally-defined item writing rules for achievement tests used to assess learning from prose material. These rules are a series of directions which tell an item writer how to rearrange segments of the instruction to obtain the items of that type. Bormuth described two general types of transformations that are possible on prose material, (a) items derived from sentences and (b) items derived from the relationships between sentences (1970, pp. 39-55). An example of sentence-derived items that test recall of prose material are those produced by the "wh-tranformation." Wh-transformation items would be written using a detailed rule summarized as follows: "Inspect all sentences in the instruction, replacing a "wh-pro" word such as who, what. or where for the appropriate part, e.g., subject, in each sentence." For instance, "The boy rode the horse" could be transformed to: "Who rode the horse?" Wh-items are particularly useful because they can be written to question each separate part of a sentence, and can be made into either multiple-choice or fill-in format.

Sentence-derived items can also be written to test comprehension of prose material through the use of paraphrasing. Anderson (1972) has emphasized the importance of paraphrasing and has defined it as the case where (a) all substantive words in a sentence are replaced and (b) the original and paraphrased sentences have equivalent meaning.

Items derived from the relationships between sentences can be produced, for example, by questioning the cause of an action or result described in prose passages. For instance, the sentences (a) Jim hurt his foot, (b) He was cleaning his gun, and (c) His gun accidentally fired, can be examined for implied causation, resulting in the question "What caused Jim's hurt foot?" (Bormuth, 1970, p. 54).

A number of question-writing methods have been proposed for testing the learning of concepts and principles (Anderson, 1972, pp. 151-159). These methods rely on an analysis of examples and non-examples of a concept or a principle, and usually go beyond the verbatim wording in prose instructional materials. Tiemann, Kroeker, and Markle (Note 1) have worked out plans for sampling examples and non-examples of concepts in both teaching and testing settings.

Finn's Extension of the Bormuth Methodology

Finn (1975) has extended the work of Bormuth by developing a question writing algorithm for prose learning. Finn's work involved a rather lengthy and detailed algorithm that originally contained 82 steps. For purposes of the present research, the original algorithm has been streamlined and the important steps in the algorithm can be described as follows:

1. Analyzing the text,
2. Selection of question words and the sentences in which they occur,

3. Preparation and analysis of the selected sentences,

4. Transformation of the sentence into a question, and

5. Generation of foils for the question.

Analyzing text. To develop questions that measure important aspects of a prose passage requires an objective analysis of the text. One approach to this has been proposed by Finn (Note 3), who uses a computer analysis of the words in a prose passage. All the words in the passage are keypunched for input to a computer program that does two major tasks: (1) counts the number of times that each word appears in the passage and (2) identifies the standard frequency index (SFI) of each word. The SFI of each word is a numerical estimate of how often the word occurs in a large corpus of American English (Carroll, Davies and Richman, 1971). A large computer tape containing the SFI of more than five million words is used to identify the SFI of each word in the passage. The computer tape used is a computerized version of the Carrol, Davies and Richman (1971) book. SFI ranges from 88.6 for the word "the" (meaning that the average American student is likely to encounter the word "the" once in every ten words of his school book reading) to 02.5 for the word "incarnation" (meaning that the average student is likely to encounter the word "incarnation" less often than once in every billion words of his school book reading).

The goal of this analysis is to identify "high information" words--words that are relatively rare in American English and occur only a single time in the passage. The sentences in which these high information words occur can then become candidates for transformation into questions which tap important information in the passage. High information words are those that are difficult for subjects to guess if they are deleted from a prose passage such as is done in a Cloze test (Culhane, 1970). Cloze tests use segments

of prose that are presented to a subject, usually with every fifth word deleted. The task for the subject is then to supply the missing words. The easiness with which a word is guessed or filled in by a subject on a Cloze test is a measure of the amount of information it provides. The task in a Cloze test is similar to the typical problem in information theory (Shannon and Weaver, 1949) in which a person is receiving a message but because of noise on the channel he is not always sure which message he hears. The information in a missing or garbled message is a function of the amount of doubt the receiver has about its identity and is related to the probability of occurrence of certain words or letters. A missing word which is highly predictable or a common word in the English language would present less information because subjects would have less doubt about the possibility that it occurs in the missing place.

Finn (Note 2) has done considerable research to show that the Cloze easiness of a word is predicted by the two important indexes derived from the computer analysis of a passage: (1) the standard frequency index (SFI) and (2) text frequency. Words that have a low SFI (infrequent in American school books) are typically high in information. However, there is one circumstance in which the information of these words is reduced in relation to a given passage. If the word is repeated frequently (i.e., it has a high text frequency), the information value of that word is diminished because subjects will supply it more often in a Cloze test following reading of the passage. In other words, repetition of words, even if they are rare in American English, lowers their information value. Therefore, Finn (Note 2) has proposed that candidates for good question words are those which are both rare in American English (have a low SFI) and are singletons in a prose passage (i.e., have a text frequency of 1).

Selection of question words and sentences. Not all parts of speech are equally good candidates for question words even though they may be high information words. Prepositions and articles are clearly not very useful. Verbs and adverbs, although important, are difficult words to remove from a sentence that is transformed into a question. For example, the sentence, "Finn echoed the concern of Bormuth" when transformed to "What did Finn do to the concern of Bormuth?" seems clumsy and seems to be a less important question than the question "Who echoed the concern of Bormuth?" After considerable work with attempts to produce questions from verbs, the authors of this report have concluded that the most promising parts of speech are adjectives and nouns.

A question word (QW) is the high-information noun or adjective in the sentence that is to be transformed. There are a few conditions under which a particular occurrence of a QW should not be used: (1) when the sentence is itself a question, (2) when the sentence gives directions to the reader ("Look at the map on p. 6"), (3) when the QW appears in a conditional clause, or (4) when the QW appears in a sentence that has already been used to generate a question.

Preparation and analysis of sentences. Once an important word has been identified for a possible question, the sentence it occurs in must be examined and prepared for transformation. The procedures suggested by Finn (1975) have been simplified by the authors of this report. The steps in this simplified procedure are described in Appendix C of this report. The important parts of this transformation are the identification and elimination of non-functional imbedded sentences, the replacement of indefinite pronouns with the phrase or word being referred to in a previous sentence, and the

actual rearrangement of words in the sentence. Examples of transformed sentences and questions are included in Appendix B of this report.

Transformation of the sentence into a question. The next step in processing the sentence is to eliminate the question word and to transform the sentence into a question. The question word and the entire noun phrase in which it occurs, is removed, and is replaced with a wh-word. Where several wordings are possible, an attempt is made to stay as close to the wording of the original sentence as possible. There are a number of further guidelines in Finn (1975) for preparing sentences for analysis such as replacing pronouns with their appropriate nouns or replacing references to previous sentences by clauses or phrases from those references.

The algorithm does not produce 100% agreement among item writers in all cases because of such things as the replacement of phrases from previous sentences and Finn (1975, pp. 357-363) discusses some of the discrepancies among item writers.

Generating foils. As is common knowledge among test developers, the writing of good foils for multiple-choice questions is challenging work. The first step in an algorithmic generation of foils is to classify the question words which are the correct answers to the question so that possible foils can be obtained from a list of words in the same classification. The most logical source of foil words would seem to be from the prose passage itself, but in some cases, there may be published lists of words such as that of Carroll, Davies and Richman (1971) that may be useful sources if the words cover the same type of subject matter.

Summary

In summary, there is considerable background work relevant to the development of a technology for generating test items for prose material.

Clearly, for research to be effective, it must be focused on a well-defined problem within the wide range of item writing methods advocated in the literature. Therefore, the current research investigates in more detail Finn's extension of the Bormuth technology for deriving items from sentences, which, as Cronbach (1970) indicated, needs to be more developed on a practical level to be useful by test developers. Bormuth and Cocks (Note 3) have commented also that Bormuth's earlier work was a "crude version" of the theory that he would like to see extended. Specifically, procedures for choosing words to delete in wh-transformations of instructional sentences and for developing foils for making these items into multiple-choice format are needed. Such a method would be widely applicable. Multiple-choice items are the most frequently used type in education and training, and the testing of recall of prose materials remains a basic element in many instructional programs.

Objectives

The purpose of the work described in the present technical report is twofold: (1) to compare various algorithmic methods for transforming instructional sentences into test questions, and (2) to examine the statistical quality of test items created by variations in algorithms. This research focuses on the multiple-choice item and its two important parts: the stem and the foils. The stem is that section of the question which requests a response from the student and the foils include the correct answer and the various wrong-answer alternatives. Specific research questions of the study are as follows:

1. What method for identifying a keyword to be selected from a prose passage is most useful? Specifically, how do "high information" words (rare singletons) compare to rare words with a high text frequency?

2. How do different item writers vary in their choice of foils and in their implementation of algorithmic item writing methods (e.g., how do their item difficulties compare)?

3. How do algorithmic methods of foil construction compare to methods in which each item writer selects the foils?

APPROACH

Materials Used in the Study

A prose passage, at the approximately high school level of reading, was selected on insect development (see Appendix A). The passage was subjected to the computer analysis suggested by Finn (Note 2). Algorithms for producing sentence-derived wh-transformations as described in the background section of this report and in Finn (1975) were applied to the passage. A set of eight parallel test forms of 20 items each were constructed. Internal-consistency reliability estimates (KR-20) averaged .65 for these test forms. Tests were given to subjects before and after they studied the prose passage. Examples of items are given in Appendix B. A total of 160 items were, therefore, the objects of analysis for this study. Several contrasting methods of algorithmic and non-algorithmic item writing were compared across four different item writers.

Selection of Question Words

All words in the passage were subjected to a computer analysis that identified the text frequency and the standard frequency index of each word in the passage. Words with a SFI (Carroll, Davies and Richman, 1971) of 60 or less were selected for possible question words in the categories of adjectives and nouns. Among all of the adjectives or nouns with a SFI of 60 or less, two major types of words were selected: (1) words having a text frequency of 1 and (2) words having the highest text frequency. As was mentioned previously in this report, the words with a text frequency of 1, which also have a SFI of 60 or less, are considered to be high information words. In the case of the words which had a high text frequency, it was necessary to select a particular occurrence of the word in the text. For

purposes of the present research an algorithm was constructed based on the choice of the first occurrence of a high text frequency word. These words will be referred to as "keywords", and high information words as "rare singleton" words.

Item Writers

Patrick Finn and three graduate students from the State University of New York at Buffalo analyzed the text and produced the questions for the research study. This group worked as a team in producing several trial questions on the prose passage in order to be sure that close agreement was made as to the classification of words and transformation of sentences. After this initial team work, the writers worked independently in producing questions.

Because all writers do not generate 100% agreement in wording of transformed sentences, the experiment examined the differences between items generated by each writer.

Development of Foils

The present research experimented with a foil writing method based on the Carroll, Davies and Richman (1971) list of words in conjunction with the prose passage on insect development (see Appendix A). The location of the question word in the Carroll list was identified and then semantically similar words in adjacent parts of the list were identified. These adjacent words, having similar SFI's were used as foils. In this particular case, the Carroll list proved to be an unacceptable source, because it produced foils which were too obviously in the wrong subject matter area, and were just plain absurd in many cases. Therefore, another method of algorithmic foil construction was developed for this experiment. The algorithm relies on

extracting foils from the prose passage itself. Two variations of the algorithm were developed: one for nouns and one for adjectives. In the case of nouns, all nouns with a standard frequency index of 60 or less were semantically classified using the method of Fredericksen (1975) as shown in Figure 1. For example, some nouns were classified as a "41" which was a concrete animate noun. For a given question word, a random sample of other nouns from the passage that were in the same semantic classification were drawn to create foils.

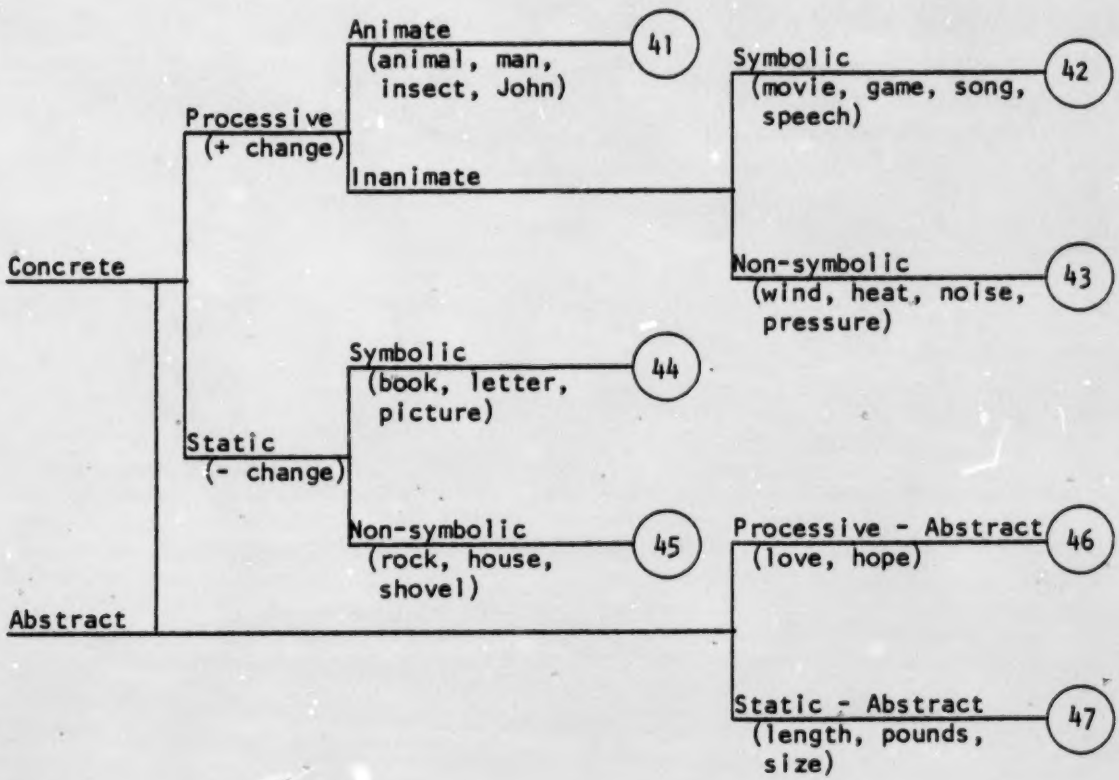


Figure 1. Fredericksen's semantic classification of nouns.

In the case of adjectives, research on semantic differential technique was used as a basis for constructing a classification of adjectives from the passage (Nunnaly, 1967. pp. 536-538). In semantic differential research there are typically three or four factors that are descriptive of words: (1) evaluation, such as "good," "bad," etc., (2) potency, such as "strong," "weak," etc., (3) activity, such as "fast," "slow," and (4) familiarity, such as "simple," "complex." These four classifications were used to group the adjectives in the passage that had SFI's less than 60. These adjectives were then subjected to an analysis of whether or not they were "familiar," as determined by the Dale-Chall List of 3,000 Familiar Words (Dale and Chall, 1948). The adjective, to be a candidate for a foil, needed to be absent from that list. The rationale for this was that there are some extremely common adjectives that would be too easy as foils.

From the insect development passage (Appendix A), there were approximately 50 adjectives which passed this screen and were qualified as possible adjectives for foil construction. A fifth classification of adjectives was used in this study. Particularly in prose material that is technically oriented, a "technical word" category of adjectives can be useful in grouping those words which are very specific to a subject matter in a particular passage. This was done with reasonably good agreement between judges. In the case of the insect development passage, technical words were those which were of scientific origin describing insect development. To build foils for adjective question words, a random sample of adjectives was drawn from the classification category of similar adjectives.

The algorithmic method of foil construction was compared, in the present research, to an informal method where each item writer chose their own foils. Even when algorithmic foils are used, the identify of the writer may show in

the choice of wording that produced the stem of the question. For this reason, the writer and foil methods are true crossed factors in this experimental design. As described above, separate methods of classifying foils were used for nouns and adjectives.

Items

Five sentences from the passage were selected for transformation into items in each of the following categories: (1) keyword adjectives which are adjectives with a SFI of 60 or less but which had a high text frequency, (2) keyword nouns which are nouns with a SFI of 60 or less but with a high text frequency, (3) rare singleton adjectives, and (4) rare singleton nouns. This represents a total of 20 sentences from the prose passage. Each sentence was transformed by each of four item writers who used two foil methods. This resulted in 20 x 4 x 2 = 160 items. Examples of each are shown in Appendix B.

Research Hypothesis

It was hypothesized that rare singleton adjectives and nouns would provide questions with the best "instructional sensitivity." Instructional sensitivity is defined as a statistically significant difference between the pretest item difficulty and the posttest item difficulty. Item difficulty is the percentage of subjects who get an item correct. The magnitude of item instructional sensitivity can be examined directly by subtracting the pretest item difficulty from the posttest item difficulty. In addition to this research hypothesis, the experiment to be described was exploratory in nature and was aimed at examining the effectiveness of the algorithm, the differences between item writers and the differences between foil construction methods in a descriptive way.

Subjects

A sample of 24 college students and 249 high school students served as subjects in the experiment. Thus, a total of 273 subjects were randomly assigned to one of eight forms of a pretest consisting of 20 items. They then studied the prose passage on insect development and were given a posttest form consisting of 20 items, different from the form given as a pretest. Each 20-item test had a balanced mixture of items developed from each algorithmic method and type of foil construction. Thus, approximately thirty subjects were randomly assigned to each test form on the pretest and the posttest.

Data Analysis

The number of subjects responding to each test form varied from 27 to 38 on the pretest and from 23 to 33 on the posttest. This provided quite stable estimates of item difficulties. A repeated-measures analysis of variance was used to compare items of each type. A 4 x 2 x 2 x 2 x 2 factorial design was used to examine differences in item difficulties between (1) the four item writers, (2) the two parts of speech (adjectives and nouns) of question words, (3) the two types of text frequencies (keyword and rare singletons), (4) the two foil types (writer's choice and algorithmic), and (5) the two test occasions (pretest and posttest).

With 160 items given on two occasions, the analysis had 320 data points, and five replications per cell. The analysis of variance was conducted on the item difficulties for items in each cell of the design.

The analysis of variance is useful for determining the "instructional sensitivity" of items. A significant main effect for the pretest-posttest factor would indicate that pretest difficulties were significantly different from posttest difficulties for all items. A significant interaction effect

involving the pretest-posttest factor would indicate that certain types of items differed in the pattern of their pretest and posttest difficulties.

RESULTS AND DISCUSSION

The results of the analysis of variance of item difficulties are presented in Table 1. The strongest effect was the main effect for test occasions indicating that, across all types of items the percentage of subjects getting pretest items correct was lower than the percentage of subjects getting posttest items correct. In other words, most items showed instructional sensitivity. Pretest item difficulties averaged 47.6% and posttest item difficulties averaged 74.4% across all items as shown in Table 2. This indicates that the subjects did learn from the passage. This also means that nearly half the subjects were able to guess the correct answer to most questions. With four-option multiple-choice items such as those used in this study, excellent items should show pretest difficulties nearer to the level of random guessing (25%).

Two important findings of this experiment were the main effect for word type (P) and the interaction of word type and the repeated measure (RP) as shown in Table 1. An inspection of average item difficulties for items constructed from each word type, as presented in Table 3, reveals these effects. Items based on question words that were nouns were significantly easier (65.6%) overall as compared to adjective items (56.5%). Table 3 reveals an ordinal interaction between the repeated measure factor (R) and part of speech (P) which indicates that noun-based items had greater instructional sensitivity than adjectives. In other words, the difference between pretest and posttest difficulties was greater for nouns than for adjectives (+29.5% vs. +24.1%).

An examination of the PS and RPS interactions, as shown in Table 4, reveals further the source of the difference between nouns and adjectives in this study. It is the keyword nouns that result in the easiest questions

Table 1

Repeated Measures Analysis of Variance on Item Difficulties of Items of Each Type

Source	df	F	P
W (Writers)	3	.25	ns*
F (Foil Type)	1	.15	ns
P (Noun vs. Adjective)	1	12.99	<.001
S (Keyword vs. Rare Singleton)	1	2.22	ns
WF	3	.54	ns
WP	3	1.66	ns
FP	1	.27	ns
WS	3	1.23	ns
FS	1	1.33	ns
PS	1	14.21	<.001
WFP	3	.11	ns
WFS	3	.34	ns
WPS	3	.25	ns
FPS	1	.13	ns
WFPS	3	.94	ns
Residual	128		
R (Pretest vs. Posttest)	1	472.03	<.001
RW	3	1.90	ns
RF	1	1.37	ns
RP	1	4.76	.03
RS	1	2.05	ns
RWF	3	1.74	ns
RWP	3	2.54	ns
RFP	1	3.04	ns
RWS	3	.43	ns
RFS	1	9.25	.003
RPS	1	20.42	<.001
RWFP	3	1.05	ns
RWFS	3	.61	ns
RWPS	3	2.63	ns
RFPS	1	.11	ns
RWFPS	3	.57	ns
Residual	128		

* ns = not significant

Table 2

Means and Standard Deviations of Item Difficulties

Type of Item		Pretest		Posttest	
		mean	s.d.	mean	s.d.
Writer (W)	1	46.8	18.3	71.6	21.4
	2	49.5	19.2	72.7	17.4
	3	46.7	20.9	75.7	20.2
	4	47.2	19.6	77.5	17.2
Foil (F)					
	Writer's Choice	46.3	20.1	74.6	19.4
	Algorithmic	48.8	18.7	74.2	18.9
Part of Speech (P)					
	Noun	50.8	20.1	80.4	14.1
	Adjective	44.3	18.2	68.4	21.5
Stem Type (S)					
	Keyword	50.4	18.4	75.4	16.4
	Rare Singleton	44.8	20.0	73.4	21.5
Test Forms	1	47.5	17.3	79.3	20.4
	2	48.6	21.4	81.9	17.6
	3	44.2	21.2	77.0	19.4
	4	51.9	18.9	74.0	21.1
	5	51.2	18.3	71.9	16.2
	6	46.9	15.3	72.4	13.4
	7	37.1	19.8	67.9	21.2
	8	53.1	20.5	70.9	21.2
All Items		47.6	19.4	74.4	19.1

Table 3

Means and Standard Deviations of Item Difficulties for the Noun and Adjective Items on Pretest and Posttest (P and RP Effects)

		Repeated Measure (R)		
Part of Speech (P)		Pretest	Posttest	Total
Noun				
	mean	50.8	80.3	65.6
	s.d.	20.1	14.1	14.7
Adjective				
	mean	44.3	68.4	56.3
	s.d.	18.2	21.5	18.1
Total				
	mean	47.6	74.4	61.0
	s.d.	19.4	19.1	17.1

- - - - - - - - -

Table 4

Means and Standard Deviations of Item Difficulties for the PS and RPS Interaction Effects

		Repeated Measure (R)		
Item Types		Pretest	Posttest	Total
Noun				
Keyword	mean	61.3	83.5	72.4
	s.d.	15.7	12.8	12.3
Rare Singleton	mean	40.4	77.3	58.8
	s.d.	18.6	14.8	13.8
Adjective				
Keyword	mean	39.4	67.4	53.4
	s.d.	14.0	15.6	13.4
Rare Singleton	mean	49.1	69.4	59.3
	s.d.	20.6	26.3	21.6
Total	mean	47.6	74.4	61.0
	s.d.	19.4	19.1	17.1

(average difficulty 72.4%), as indicated by a Newman-Kuels a posteriori analysis of the means in Table 4 (PS interaction). Keyword nouns are nouns from the passage having a SFI of 60 or less but which have a high text frequency. The first occurrence of the keyword in the text was used to create the item. An examination of the sentences in which these keyword nouns appeared provides a possible explanation for the results. These sentences are typically introductory sentences which are very general and which address the main topics of the entire passage. For example, in the passage on insect development, the keyword noun "insects" appears in the very first sentence of the passage and is a very general statement--"The life of most insects is short but active." Questions derived from this type of sentence seem to be invariably a common knowledge type of item which students can answer without having read the prose passage. Also, keyword nouns were relatively easy for subjects to recall on the posttest (average item difficulty of 83.5%). This may be due to the fact that these keywords were mentioned several times in the passage. Easiness of these items on the posttest is in line with Finn's (Note 2) hypothesis that the information content of these rare words is reduced by their high text frequency.

Finn's hypothesis concerning reduced information content of keywords would seem to be inconsistent with the fact that keyword adjectives produced the most difficult items (53.4% as shown in Table 4). However, an examination of each of the keywords provides a rationale for this inconsistency. The keywords and their text frequency are given in Table 5.

Table 5 clearly shows that the keyword adjectives had much lower text frequency than keyword nouns. Thus, Finn's hypothesis applies in that higher text frequency was related to the easiness of items constructed from keywords. With text frequencies of 2 or 3, the keyword adjectives were very close to

being rare singletons. Keyword nouns included "insect" and its plural "insects," having a total of 28 occurrences in the text. Also, plurals of "egg" and "adult" appeared 15 and 6 times, respectively, in the passage, to further increase the familiarity of these words to the reader

Table 5

Keywords in the Insect Development Passage and Their Text Frequency

Keyword Nouns	Text Frequency	Keyword Adjectives	Text Frequency
insect	8	immature	3
insects	20	incomplete	2
metamorphosis	9	nymphal	2
egg	8	aquatic	2
adult	8	distinctive	2

The rare singleton nouns (instars, cicadas, silverfish, wasps and appetites) showed a good pattern of pretest and posttest difficulties. They had the highest average instructional sensitivity (40.4% to 77.3%), a difference of 36.9% as shown in Table 4. The rare singleton adjectives (plant-feeding, pupal, spine-like, self-made and worm-like), compared to the rare singleton nouns, were somewhat easier on the pretest and more difficult on the posttest.

As shown in Table 1, there was no main effect for writers or foil type nor was there a significant two-way interaction between writers and foil. This result is somewhat surprising in that different writers would be expected to write easier or harder items when they had the latitude to choose their own foils.

Table 1 showed one interaction (RFS) involving foil type. The means and standard deviations of item difficulties for that interaction are presented in Table 6. All of the posttest means in this table are very similar. A Newman-Kuels *a posteriori* test of the differences between pretest item

difficulties in this interaction was calculated. This revealed that among the items with "writer's choice" foils, the rare-singleton items were more difficult on the pretest (40.2%) than the keyword items (52.5%).

Table 6

Means and Standard Deviations for the RFS Interaction Effect

Item Type		Pretest	Posttest	Total
Writer's Choice				
Rare Singleton	mean	40.2	74.0	57.1
	s.d.	20.2	21.4	17.9
Keyword	mean	52.5	75.3	68.9
	s.d.	18.2	17.5	16.3
Algorithmic				
Rare Singleton	mean	49.3	72.7	61.0
	s.d.	19.0	21.9	18.1
Keyword	mean	48.2	75.6	61.9
	s.d.	18.6	15.4	15.8
Total	mean	47.6	74.4	61.0
	s.d.	19.4	19.1	17.1

Variance Between Writers

Another phase of data analyses conducted for this study was the examination of the variability of item difficulties across item writers. One of the major research questions of this study was, "Do the difficulties of items constructed with the 'writer's choice' foils vary more widely across writers than the difficulties of items constructed with the algorithmic foils?". The expectation was that some writers would choose very difficult foils for a given transformed sentence and other writers would choose very easy foils for that same sentence. In contrast, the algorithmic foils which were chosen at random from matched groups of similar words from the passage,

would seem to be free of any item-writer bias, and, hence, less variable in their effects on item difficulty.

In examining the variability across writers, the focus was on each sentence that was to be transformed by each writer. There were a total of 20 sentences taken from the insect development passage. For example, each of the four item writers produced items from the sentence, "In most cases each egg produces a single immature insect," for which the keyword adjective was "immature" (see Appendix B for the complete example of this item). It was, therefore, possible to identify four item difficulties for a given combination of sentence and foil technique. For instance, for the sentence containing the word "immature," each writer generated an item using the "writer's choice" foil method, and this resulted in pretest difficulties of 38%, 65%, 52%, and 37%. Posttest item difficulties for these items were 67%, 63%, 74%, and 52%, in that order. The pretest and posttest variabilities could then be calculated across these item difficulties.

After all variances of item difficulties across writers were calculated, they were subjected to a repeated measures analysis of variance in which the dependent variables were the natural logarithms of the variances (Scheffé, 1959, p. 83). The design for this analysis was 2 x 2 x 2 x 2 with the following factors: (1) foil type (writer's choice vs. algorithmic), (2) part of speech (noun vs. adjective), (3) stem type (keyword vs. rare singleton question word), and (4) the repeated measure (pretest vs. posttest).

The surprising result of this analysis was that there were no significant main effects or interactions. For example, even though the writer's-choice foil method had an average variability of 115.31% (standard deviation of 10.74%) and the algorithmic foil method had a variability of 73.97% (s.d. of 8.60%), this difference was not statistically significant. The variabilities of item types are shown in Table 7.

One important limitation of the present study that should be mentioned is that only four item writers were employed. Calculation of variabilities across only four writers is clearly susceptible to the influence of any one of the four item difficulties. With a larger sample of writers, the effects may have been more clearly detectible.

Table 7

Variabilities and Standard Deviations of Item Difficulties

Item Types		Pretest	Posttest	Total
Foil Type				
Writer's Choice	var.	131.39	101.21	115.31
	s.d.	11.46	10.06	10.74
Algorithmic	var.	69.94	78.24	73.97
	s.d.	8.36	8.85	8.60
Part of Speech				
Noun	var.	94.45	85.63	89.93
	s.d.	9.72	9.25	9.48
Adjective	var.	97.30	92.47	94.85
	s.d.	9.86	9.62	9.74
Item Type				
Keyword	var.	102.19	87.06	94.32
	s.d.	10.11	9.33	9.71
Rare Singleton	var.	89.93	90.95	90.44
	s.d.	9.48	9.54	9.51

CONCLUSIONS

An important finding of this study is that no significant difference between item writers was found in the analysis of item difficulties. In contrast, the study by Roid and Haladyna (1978, in press) found significant differences in item difficulties between item writers who wrote from learning objectives or loosely-defined rules for transforming sentences. In other words, the methodology of transforming sentences into questions as implemented in this study apparently reduces the item bias created by differences between item writers.

The concept of identifying high information words from prose passages for use in selecting sentences to be transformed into questions is apparently a workable methodology. Words which have a SFI of 60 or less (they are relatively rare in American textbooks) and yet occur only once or twice in a prose passage are apparently good candidates for questions that test learning from prose. In contrast, nouns which are rare in American textbooks yet highly frequent in a prose passage are apparently not very good candidates for question words as they are too simple even on a pretest. This is particularly true, apparently, if the first occurrence of the noun in the passage is used.

Algorithmic methods of constructing foils for multiple choice versions of sentence-derived questions as implemented in the present study were shown to be feasible. One disadvantage of these sentence-transformation methods may be that they are somewhat easy on a pretest. However, they still appear to generate items that show significant instructional sensitivity--a shift in difficulty from pretest to posttest when instruction is given between testing sessions. Algorithmically derived foils for items functioned approximately the same as foils produced by item writers.

The present study provides evidence to support the theory of processing in reading proposed by Finn (Note 2). Rare singleton nouns, as compared to keyword nouns, were shown to be much more difficult for subjects to identify as answer words. In other words, rare singletons are truly high information words. The present study further suggests that rare words that occur on two or three occasions in a passage may have the same attributes as rare singletons.

RECOMMENDATIONS

1. Algorithmic generation of multiple choice test questions by transforming sentences from instructional materials should be considered feasible and ready for further refinement and application in a variety of subject matter areas.

2. The concept of "high information" words should be considered a feasible and useful basis for selecting important sentences from prose passages for transformation into questions which measure reading comprehension.

3. Keyword nouns, which are relatively rare in American textbooks but occur very frequently in a prose passage, should not be used as question words, particularly, if the sentence in which they first occur is used to form the question.

REFERENCES

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42. 145-170.

Bormuth, J. R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Bloom, B. S. Learning for mastery. Evaluation Comment, UCLA, Vol. 1, No. 2, May, 1968.

Carroll, J. B., Davies, P., & Richman, B. Word frequency book. Boston: Houghton-Mifflin, 1971.

Cronbach, L. J. John R. Bormuth: On the theory of achievement test items. Psychometrika, 1970, 35, 509-511. (Book Review)

Culhane, J. W. CLOZE procedures and comprehension. The Reading Teacher, 1970, 23, 410-413.

Dale, E., & Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-28.

Finn, P. J. A question writing algorithm. Journal of Reading Behavior, 1975, 4, 341-367.

Fredericksen, C. H. Representing logical and semantic structure of knowledge acquired from discourse. Cognitive Psychology, 1975, 7, 371-458.

Hively, W. Introduction to domain-referenced testing. Educational Technology, 1974, 14, 5-10.

Millman, J. Criterion-referenced measurement. In Popham, W. J. (Ed.) Evaluation in education: Current applications. Berkeley, California: McCutchan Publishing Company, 1974.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1967.

Osburn, H. G. Item sampling for achievement testing. Educational and Psychological Measurement, 1968, 28, 95-104.

Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.

Roid, G. H., & Haladyna, T. A comparison of objective-based and modified-Bormuth item writing techniques. Educational and Psychological Measurement, Spring, 1978, in press.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Shannon, C. E., & Weaver, W. The mathematical theory of communication. Urbana: University of Illinois Press, 1949.

Swezey, R. W., & Pearlstein, R. B. Developing criterion-referenced tests. Reston, Virginia: Applied Science Associates, 1974.

REFERENCE NOTES

1. Tiemann, P., Kroeker, L. P., & Markle, S. M. Teaching verbally-mediated coordinate concepts in an on-going college course. Paper presented at the meetings of the American Educational Research Association, New York, April, 1977.

2. Finn, P. J. Word frequency, information theory and cloze performance: A lexical-marker, transfer-feature theory of processing in reading. Unpublished paper, State University of New York at Buffalo, School of Education, 1977. (Submitted for publication in Reading Research Quarterly.

3. Bormuth, J. R., & Cocks, P. Rules for classifying and scoring responses to wh-questions. Paper presented at the meetings of the American Educational Research Association, Washington, D.C., March, 1975.

APPENDIX A

"INSECT DEVELOPMENT"

THE PROSE PASSAGE USED IN THE EXPERIMENT

4. INSECT DEVELOPMENT

The life of most insects is short but active. Very few insects have a life-span of more than a year. By a life-span we mean the time from when the egg is laid to when the fully developed adult dies. Let's look at what happens during this period.

All insects develop from eggs. In most cases these eggs hatch outside the body of the female. In the few cases in which the eggs hatch inside the female the young are born "alive." These insects, such as the aphids, are said to be viviparous. (vy-vip'-ah-rus).

Insects that hatch from eggs after they have been laid are said to be oviparous (oh-vip'-ah-rus). Most insects are oviparous. In most cases each egg produces a single immature insect. However, in certain species of parasitic wasps (encyrtids), the egg may produce two or more young.

Most insect eggs are very distinctive. The size, shape, or color of the egg is different, in most cases, for each species of insect. This enables a person who has made a study of these eggs to identify the insect that laid them almost as easily as if he had seen the adult.

Most insect eggs are laid in a place that will provide either protection or food for the young. Protection is especially important to those insects that overwinter in the egg stage. Overwintering means that the adult insect lays its eggs in the late summer or early fall. The eggs then are dormant until the next spring when they hatch. Most of the adults of these species are killed by the first frost. However, the hatching of these eggs in the spring produces new individuals to carry on the species.

Most plant-feeding insects instinctively lay their eggs on plants that the young feed on. This increases the immature insects' chances of survival. If this field of investigation interests you, the study and photography of insect eggs might make a good project.

After reaching the proper stage of development, the egg will hatch. The young insect can use a number of ways to get out of the egg. Some insects

chew their way out. Others have special spinelike structures, called egg-bursters, which cut through the shell. There are some eggs which have special weak spots in them. The young insect escapes from these either by wriggling or by taking in air and bursting the shell with internal pressure.

After the Egg

After hatching, all insects, except the most primitive, go through a series of steps in development. These steps are called *metamorphosis*. The word metamorphosis comes from two Greek words: meta, meaning to change, and morpho, meaning form. Therefore, metamorphosis means a change in form. This change in form occurs in two different ways. These two ways are called complete and incomplete metamorphosis. The most primitive insects, such as the silverfish, do not go through metamorphosis. When they hatch they look like their parents in every way except that they are smaller. Their development consists of growing larger and becoming able to reproduce.

Incomplete Metamorphosis

Insects which show this type of metamorphosis have young which look very much like the adults of the species. These immature insects are called nymphs. With the exception of some aquatic species, the principal differences between the nymphs and adults are in size and the presence of wings (see illustration at the right).

Now think back to the description of the phylum to which insects belong, *Arthropoda*. Remember, one of the characteristics of these animals is a hard outer covering called an *exoskeleton*. The exoskeleton is made of a nonliving substance called chitin (ki'-tin). Chitin is hard and stiff and has very little "stretch." Inside the exoskeleton there is very little room for growth.

In order to grow, the nymph must escape this self-made prison. It does this by secreting a new exoskeleton under the old one. When this new skin is complete the old skeleton splits down the

back and the insect walks away and leaves it behind. You have probably seen some of these discarded skins, called casts, on tree trunks.

For a time after the insect discards its old skin, the new exoskeleton is soft. This allows the exoskeleton to expand and make room for further growth.

Each of the periods between molts is called an *instar*. Some nymphs go through as many as eight or more instars before emerging as adults.

Aquatic species that undergo incomplete metamorphosis must go through one more step in development. As nymphs they breathe by means of gills. These gills must be replaced by air-breathing organs in the adult stage. This is done in the last nymphal instar. When it is time for the adult to emerge, the nymph rises to the surface and molts. The fully developed adult steps out of the final nymphal skin with fully developed organs for breathing air.

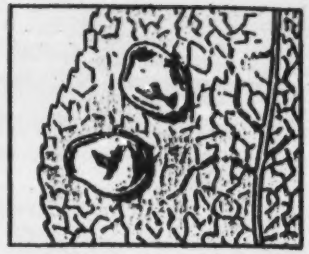
Complete Metamorphosis

This is the type of metamorphosis that most people are familiar with. Butterflies and moths have complete metamorphosis. There are four distinct stages: egg, larva, pupa, and adult. Since the adult's main activity is producing eggs, and I'm sure you know what these are, we will spend our time studying the larva and pupa.

The larvae's main job in life is to eat and grow. They have huge appetites. Larvae are very different from the adults. They do not have compound eyes, wings, and usually have chewing mouth parts even in those orders where the adults have sucking mouth parts.

A larva may continue to eat and grow all summer. As cold weather approaches, it may build a cocoon and pass into the pupal stage.

Most of these insects pass the winter inside the cocoon. Because no activity is visible at this time, the pupa has been falsely called a "resting stage." Actually a great deal of activity is going on. The wormlike larva is changing into a fully developed adult. When the weather is warm again, this adult emerges from the cocoon, mates, lays eggs, and starts the whole process over again.



Cecropia eggs

—H. B. Kane

Cecropia larva

—G. A. Smith

Spinning a cocoon

—G. McCormack

The pupa stage

—G. McCormack

An emerging adult

—G. McCormack

An adult cecropia moth

DEVELOPMENT OF A MOTH

Let's Get Together

Most insects reproduce sexually. This means that, to have eggs that will hatch, a male and a female of the species must mate. The question is: How do they find each other?

It has been known for years that some of the sounds made by crickets and cicadas were a type of mating call. It is easy to see how these insects get together. But what about the insects that do not make noise; butterflies, for instance?

It has been discovered that the females of these species give off a distinctive odor. This odor is detectable by male insects over great distances. The male follows this scent trail back to the female.

This brings to mind an interesting experiment you might try. A friend of mine once caught a recently emerged female Promethea moth. He put the female in a screen cage and set it outside his window. In less than two hours there were more than twenty males hanging on the outside of the cage. Why don't you try this with other kinds of insects? It would make a great science project.

Science has used the discovery of these odors to help eliminate undesirable insects. It was found that female cockroaches gave off an attractive (to male cockroaches) odor. Scientists have been able to reproduce this scent and have used it to attract males to traps.

Exercises

How Well Did You Read?

1. Name and describe the three types of development insects can go through.
2. What advantage is there in insect eggs being laid on certain plants?
3. What is metamorphosis? What are the differences between complete and incomplete metamorphosis?
4. What processes take place during the growth of insects?
5. Can you think of any advantages to some insects in being born "alive"?

Read A Little More

1. Lemmon, R. S., *All About Moths and Butterflies*, New York: Random House, 1956.

APPENDIX B

EXAMPLES OF ITEMS PRODUCED FROM THE PASSAGE "INSECT DEVELOPMENT"

EXAMPLES OF ITEMS PRODUCED FROM THE PASSAGE "INSECT DEVELOPMENT"

1. Keyword Adjectives

WORD: Immature

TEXT SENTENCE(S): In most cases each egg produces a single immature insect.

ITEM (ITEM WRITER #4, "WRITER'S CHOICE" FOILS):

What does each egg produce in most cases? A single

a. oviparous insect
b. nymphal insect
c. mature insect
*d. immature insect

ALGORITHMIC FOILS:

a. complete insect
b. distinct insect
*c. immature insect
d. incomplete insect

2. Rare Singleton Adjectives

WORD: Pupal

TEXT SENTENCE(S): A larva may continue to eat and grow all summer. As cold weather approaches, it may build a cocoon and pass into the pupal stage.

ITEM (ITEM WRITER #3, "WRITER'S CHOICE" FOILS):

As cold weather approaches, what may a larva do? Build a cocoon and pass into the

*a. pupal stage.
b. dormant stage.
c. hibernation stage.
d. resting stage.

ALGORITHMIC FOILS:

*a. pupal stage.
b. nymphal stage.
c. parasitic stage.
d. insect stage.

3. Keyword Nouns

WORD: Metamorphosis

TEXT SENTENCE(S): After hatching, all insects, except the most primitive, go through a series of steps in development. These steps are called metamorphosis.

ITEM (ITEM WRITER #1, "WRITER'S CHOICE" FOILS):

What are a series of steps in development called?

*a. metamorphosis
b. reproduction
c. larvae
d. changes

ITEM (ITEM WRITER #3, "WRITER'S CHOICE FOILS):

What are the series of steps in insect development called?

a. maturation
*b. metamorphosis
c. symbiosis
d. meitosis

ALGORITHMIC FOILS:

a. growths
*b. metamorphosis
c. types
d. activities

4. Rare Singleton Nouns

WORD: Silverfish

TEXT SENTENCE(S): The most primitive insects, such as the silverfish, do not go through metamorphosis.

ITEM (ITEM WRITER #2, "WRITER'S CHOICE" FOILS):

What insects do not go through metamorphosis? The primitive, such as

a. eggs
*b. silverfish
c. chitin
d. butterflies

ALGORITHMIC FOILS:

*a. silverfish
b. females
c. individuals
d. wasps

APPENDIX C

STEPS IN TRANSFORMING SENTENCES INTO TEST QUESTIONS

STEPS IN TRANSFORMING SENTENCES INTO TEST QUESTIONS

1. Take the first question word (QW) under consideration and identify the sentence it occurs in, using text punctuation for guidance.
2. Go on to the next QW (do not use the QW now under consideration) if:
   a. Sentence is itself a question.
   b. Sentence is directions to reader. (Look at map on page 2.)
   c. QW appears in a conditional clause.
   d. QW appears in a sentence that has already been used to generate a question. (Choose alternate keyword or go to next singleton.)
3. Copy out sentence using text punctuation for guidance.
4. Identify all indefinite pronouns such as "this" or "it" that refer to a previous sentence and insert the word or phrase from the previous sentence in place of the pronoun.
5. Identify all imbedded sentences.
6. Eliminate imbedded sentences that do not contain QW using the following criteria:
   a. Does this sentence limit the meaning of the larger sentence? If yes, keep it.
   b. Would elimination of this sentence make the larger sentence ungrammatical? If yes, keep it.
   c. Would elimination of this sentence make the larger sentence an incomplete thought? If yes, keep it.
   d. If answer to questions a, b, and c are all <u>no</u>, eliminate the imbedded sentence under consideration.
7. Eliminate QW and replace it with an appropriate wh-wor
   a. Eliminate the entire noun phrase.
   b. If QW is part of a compound construction, eliminate the entire construction.

8. Generate a question by placing an appropriate word such as "what" as the first word in the question.

9. Where several wordings are possible, stay as close to the wording of the original sentence as possible.

10. Eliminate all redundancy in the foils by stating all non-QW elements in the entire noun phrase as a "fill-in" statement prior to the foils, e.g., the sentence, "Each egg produces a single <u>immature</u> insect," would be transformed to "What does each egg produce? A single _____ insect."